2000

# Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations

J. A. Labate
*Cornell University*

K. R. Lamkey
*U.S. Department of Agriculture*, krlamkey@iastate.edu

M. Lee
*Iowa State University*, mlee@iastate.edu

W. Woodman
*Iowa State University*

# Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations

**Abstract**
Because maize (*Zea mays*) is an annual species those working with it must frequently make crosses to preserve and periodically maintain populations. Random mating is performed either using hand-pollination techniques or in wind-pollinated isolated blocks. Eighty-two restriction fragment length polymorphism (RFLP) markers were used to examine samples of random mated, hand-pollinated BSSS(R) and BSCB1(R) maize populations to find out whether their genotypic proportions conformed to predicted outcomes of random mating. The majority of loci conformed to expectations for Hardy-Weinberg equilibrium (HWE). Excess homozygosity was observed at 87% of the loci where the null hypothesis of HWE was rejected. For pairs of polymorphic loci, linkage equilibrium was observed in the BSSS(R) and BSCB1(R) progenitor populations (fewer than 5% of all tests rejected the null hypothesis of equilibrium at the $P \leq 0.05$ significance level). The BSSS(R)CO, BSCB1(R)CO and BSCB1(R)C12 populations showed slight increases in the proportion of pairs of loci in linkage disequilibrium compared to the progenitors (approximately 8.4% of all pairs of loci rejected the null hypothesis at the $P \leq 0.05$ significance level). BSSS(R)C12 was an extreme outlier with 25.0% of all pairs of polymorphic loci displaying significant ($P \leq 0.05$) linkage disequilibrium. This result was likely caused by the artificial grouping of three BSSS(R)CO plants with 97 BSSS(R)C12 plants during sampling. Results from principal components analysis of all individuals based on RFLP alleles supported this interpretation. Overall, most of the observed deviations from equilibrium were likely to have been caused by positive assortative mating in the case of HWE, and natural selection for epistatic effects between unlinked loci in the case of linkage disequilibrium.

**Keywords**
RFLP markers, Principal component analysis, Genotypic frequencies, Quantitative genetics, Zea mays L.

**Disciplines**
Agricultural Science | Agronomy and Crop Sciences | Plant Breeding and Genetics

**Comments**
This article is published as Labate, J. A., K. R. Lamkey, M. Lee, and W. Woodman. 2000. Hardy-Weinberg and Linkage Equilibrium Estimates in the BSSS and BSCB1 Random Mated Populations. Maydica 45:243-256.

Maydica 45 (2000): 243-255

# HARDY-WEINBERG AND LINKAGE EQUILIBRIUM ESTIMATES IN THE BSSS AND BSCB1 RANDOM MATED POPULATIONS[1]

**J.A. Labate[2], K.R. Lamkey[3]*, M. Lee[4], W. Woodman[4]**

[2] *Institute for Genomic Diversity, Cornell University, Ithaca, NY, 14853*
[3] *USDA-ARS, Dept. of Agronomy, Iowa State Univ., Ames, IA 50011*
[4] *Dept. of Agronomy, Iowa State Univ., Ames, IA 50011*

Received July 28, 2000

**ABSTRACT** - Because maize (*Zea mays* L.) is an annual species those working with it must frequently make crosses to preserve and periodically maintain populations. Random mating is performed either using hand-pollination techniques or in wind-pollinated isolated blocks. Eighty-two restriction fragment length polymorphism (RFLP) markers were used to examine samples of random mated, hand-pollinated BSSS(R) and BSCB1(R) maize populations to find out whether their genotypic proportions conformed to predicted outcomes of random mating. The majority of loci conformed to expectations for Hardy-Weinberg equilibrium (HWE). Excess homozygosity was observed at 87% of the loci where the null hypothesis of HWE was rejected. For pairs of polymorphic loci, linkage equilibrium was observed in the BSSS(R) and BSCB1(R) progenitor populations (fewer than 5% of all tests rejected the null hypothesis of equilibrium at the $P \leq 0.05$ significance level). The BSSS(R)C0, BSCB1(R)C0 and BSCB1(R)C12 populations showed slight increases in the proportion of pairs of loci in linkage disequilibrium compared to the progenitors (approximately 8.4% of all pairs of loci rejected the null hypothesis at the $P \leq 0.05$ significance level). BSSS(R)C12 was an extreme outlier with 25.0% of all pairs of polymorphic loci displaying significant ($P \leq 0.05$) linkage disequilibrium. This result was likely caused by the artificial grouping of three BSSS(R)C0 plants with 97 BSSS(R)C12 plants during sampling. Results from principal components analysis of all individuals based on RFLP alleles supported this interpretation. Overall, most of the observed deviations from equilibrium were likely to have been caused by positive assortative mating in the case of HWE, and natural selection for epistatic effects between unlinked loci in the case of linkage disequilirium.

KEY WORDS: RFLP markers; Principal component analysis; Genotypic frequencies; Quantitative genetics; *Zea mays* L..

## INTRODUCTION

Maize populations are random mated to form new populations, to maintain seed of existing populations, and to multiply seed of existing populations for experimental use. These three applications may have slightly different goals concerning the genetic structure of the population formed by random mating. In all three circumstances we expect populations that are in Hardy-Weinberg proportions (HWP) for single loci, i.e., genotype frequencies can be predicted by population allele frequencies. If a breeder is forming a new population, the usual concern is to ensure equal representation of all parents in the final population. Researchers who are generating seed for experimental use are interested in sampling enough unique/unrelated plants to obtain accurate estimates of various genetic parameters. Germplasm managers who are maintaining seed of populations as genetic repositories are interested in maintaining rare alleles (CROSSA *et al.*, 1993; 1994). In the former two cases rare alleles are not critical to maintain because their contribution to the total genetic variance is usually negligible.

Linkage equilibrium in a population, i.e., the random association of alleles at different loci with each other, can be disrupted by such factors as admixture of populations, genetic drift, segregation of chromosomal inversion polymorphisms, or selection for epistatic interactions. Hardy-Weinberg and linkage equilibrium are basic assumptions of many quantitative genetics models. Maize has been one of the most extensively studied plant species from a quantitative genetics perspective (HALLAUER and MIRANDA, 1981) but very little is known about whether

[1] Joint contribution from the Corn Insects and Crop Genetics Research Unit, USDA, Agricultural Research Service, Dept. of Agronomy, Iowa State Univ. and Journal Paper No. J-18766 of the Iowa Agric. and Home Economics Exp. Stn. Project No. 3495. Received.

* For correspondence (fax +1 515 294 9359; e-mail: krlamkey@iastate.edu).

the majority of studied maize populations are in Hardy-Weinberg or linkage equilibrium.

There are two broad categories of random mating methods in maize: open-pollination in isolation and artificial pollination by hand. In theory a truly random mating system will include selfing of the individuals being random mated (FALCONER and MACKAY, 1996), but in practice this is nearly impossible to obtain. With both methods the goals are to equalize genetic contributions of each parent and to maintain large effective population sizes (CROSSA, 1989). CROSSA (1989) outlined an optimum method for random mating maize populations by artificial pollination. The number of seeds kept per harvested ear is critical in determining the effective size of the population as well as its usefulness when being sampled for breeding purposes (CROSSA, 1989). Open-pollination and hand-pollination will give the same results if properly performed.

In a reciprocal recurrent selection program the final phase of each cycle of selection is recombination of the selected progenies to form a new population (HALLAUER, 1985). It has been estimated that four to five generations of random mating will result in a population approaching linkage equilibrium (HANSON, 1959), but generally only one or two generations are advanced per cycle so as not to substantially increase cycle time (HALLAUER, 1985).

The purpose of this study was to measure Hardy-Weinberg proportions at single loci, and linkage disequilibrium between pairs of loci, in random mated, hand-pollinated BSSS(R) and BSCB1(R) populations using RFLP markers. Sampled Cycle 0 plants had undergone several decades of random mating with large effective population sizes. Cycle 12 plants were sampled from Syn-3 (Synthetic-3) populations, which originated from the intermating of 20 selected lines that subsequently experienced two generations of random mating (sampling 350 ears each generation). We expect to observe Hardy-Weinberg proportions in these maize populations because these are produced from a single generation of random mating with nonoverlapping generations. It is less obvious what to predict in terms of linkage equilibrium in the populations. Ignoring selection and drift that could obviously have contributed during their history, linkage disequilibrium can be created if alleles are not randomly distributed among gametes during the founding of a population. The decay of linkage disequilibrium thus generated depends on recombination rates between loci.

## MATERIALS AND METHODS

### *Sampling of Populations*

We have genotyped samples from three populations within BSSS(R) and BSCB1(R), representing three different stages of their history (see LABATE *et al.*, 1997 for complete details). BSSS and BSCB1 synthetic populations trace back to 16 and 12 inbred lines, respectively. These collections of inbred lines are herein referred to as progenitor (P) populations. Cycle 0 (C0) populations were formed by several generations of random mating bulked seed obtained from a series of crosses between progenitor-inbred lines. These BSSS(R)C0 and BSCB1(R)C0 populations were the starting material for reciprocal recurrent selection (RRS). They had been periodically maintained as large, random mated populations (recombining 300-500 plants each generation) for several decades before samples were taken for our study, although there may have been some undocumented genetic bottlenecks (fewer than 300 plants were used) in their distant past, pre-1970. Samples from both populations were also genotyped after 12 cycles of RRS (C12). The Cycle 12 populations had undergone two generations of random mating prior to sampling.

The molecular markers used were 82 nuclear genomic restriction fragment length polymorphism (RFLP) loci randomly distributed across all 20 chromosomal arms. Probe names are given in Fig. 3. The probes were chosen for their high levels of polymorphism and extensive coverage of the genome. One hundred individuals from each Cycle 0 and Cycle 12 population were chosen at random for genotyping, as well as single individuals from each of 28 progenitor inbred lines (two of the BSSS progenitor inbred lines had been lost; however, the two parental lines of one of these were included). Each of the 82 RFLP probes was considered to be a single locus and variants at each locus were assumed to be allelic.

### *Hardy-Weinberg Tests*

A test of the null hypothesis of random union of gametes was performed for all polymorphic (most common allele was not fixed) loci in the Cycle 0 and Cycle 12 populations. The population genetic software package GENEPOP (RAYMOND and ROUSSET, 1995) was used to calculate exact probability of type I error by complete enumeration for loci with less than five alleles (LOUIS and DEMPSTER, 1987). For loci with five or more alleles, GENEPOP used a Markov chain method (GUO and THOMPSON, 1992) to estimate an unbiased exact probability of type I error and a corresponding standard error. The analyses were repeated if necessary using an increased number of "batches" (see GENEPOP manual) until all standard errors were less than 0.01.

For loci at which the null hypothesis was rejected ($P \leq 0.05$), the proportion of excess homozygosity was estimated as $(H_O - H_e)/n$, where $H_O$ was the observed number of heterozygotes, $H_e$ was the expected number of heterozygotes under Hardy-Weinberg proportions, and $n$ was the sample size.

### *Two-locus Disequilibrium Tests*

GENEPOP (RAYMOND and ROUSSET, 1995) was used to test for genotypic independence between pairs of loci. Genotypic disequilibrium was estimated for all possible (maximum 3,321) pairs of loci in progenitor, Cycle 0, and Cycle 12 populations. The exact Fisher test on a row by column contingency table of genotypic frequencies was performed using a Markov chain method that estimated the exact probability of type I error without bias. The null hypothesis was that the rows and columns were indepen-

TABLE 1 - *Number of pairs of loci in significant linkage disequilibrium in two maize populations undergoing reciprocal recurrent selection.*

| Population | Significance level | | | | | | |
|---|---|---|---|---|---|---|---|
| | $P < 0.001$ | | $0.001 < P \leq 0.01$ | | $0.001 < P \leq 0.05$ | | |
| | Number of pairs | Number of loci involved | Number of pairs | Number of loci involved | Number of pairs | Number of loci involved | Total number of pairs[a] |
| BSSS(R)P | 1 | 2 | 14 | 28 | 81 | 70 | 3,240 |
| BSCB1(R)P | 2 | 4 | 19 | 34 | 85 | 68 | 3,321 |
| BSSS(R)C0 | 31 | 43 | 51 | 66 | 220 | 80 | 3,160 |
| BSCB1(R)C0 | 29 | 42 | 65 | 72 | 225 | 82 | 3,321 |
| BSSS(R)C12 | 190 | 49 | 164 | 65 | 398 | 77 | 3,003 |
| BSCB1(R)C12 | 14 | 23 | 22 | 41 | 99 | 65 | 2,211 |

[a] Total number of pairs = $n(n-1)/2$, for $n$ polymorphic loci.

dent. To process the GENEPOP output, PASCAL programs were used to identify pairs of loci with $P$ values falling within certain ranges (Table 1), and to create matrices summarizing all pairwise results (Fig. 3 and Fig. 4).

### *Principal Components Analysis*

The software program NTSYS-pc (Rohlf, 1994) was used to perform a principal components analysis (PCA) on the 428 individuals sampled from BSSS(R) and BSCB1(R) populations based on genotypes for all loci. This was done in order to examine clustering of individuals based on their genotypes. Discrete subpopulations can be evident within a population when there is extensive linkage disequilibrium. The data were coded as 0 or 1 (absence or presence of an allele) for all 391 alleles at the 82 loci. A correlation matrix between individuals was computed using the simple matching coefficient ($S_{SM}$) (Sneath and Sokal, 1973, p. 132). The first three principal components were extracted from the correlation matrix (using EIGEN) and the individuals were projected onto the PCA axes (using PROJ).

## RESULTS

### *Single-locus Equilibrium*

The numbers of loci that were not in Hardy-Weinberg equilibrium are summarized in Fig. 1. There were 80 polymorphic (unfixed) loci in BSSS(R)C0, 78 in BSSS(R)C12, 82 in BSCB1(R)C0, and 67 in BSCB1(R)C12. The total number of statistically significant ($P \leq 0.05$) Hardy-Weinberg tests was 83. All four populations showed similar numbers of loci in significant disequilibrium. The Cycle 12 populations had fewer loci in disequilibrium than the Cycle 0 populations, but this may have been because there were fewer polymorphic loci. The number of instances of disequilibrium did not vary widely among the 10 chromosomes. It ranged from a total of three cases (chromosome three), to 12 cases (chromosome one). Chromosome three had the fewest RFLP markers (five); the other chromosomes carried between seven and 12 marker loci each.

The most important observation concerned the direction of the deviations from Hardy-Weinberg expectations. Excess homozygosity was observed in 87% of the cases of single-locus disequilibrium, and the degree of excess homozygosity followed a normal distribution (Fig. 2) (Shapiro-Wilk W = 0.9855, $P < 0.8330$) (Shapiro and Wilk, 1965). The relative number of cases of excess homozygosity versus excess heterozygosity was similar in all four populations, approximately 6:1 (Fig. 1). Mean proportion of excess homozygosity in the significant tests over all four populations (Fig. 2) was 12.2%. If there were multiple alleles at a locus, in general, homozygous excess was observed in every allelic class (data not shown). This implied the union of similar gametes at individual loci. The fraction of loci displaying excess homozygosity within a population ranged from 12/78 loci (15.4%) in BSSS(R)C12 to 26/82 loci (31.7%) in BSCB1(R)C0. There was no tendency for an increase in excess homozygosity in Cycle 12 compared to Cycle 0. The mean proportion of excess homozygosity for each population was: 13.8 ± 2.42% for BSSS(R)C0, 13.2 ± 3.53% for BSSS(R)C12, 13.1 ± 2.05% for BSCB1(R)C0, and 8.1 ± 3.51% for BSCB1(R)C12.

The inbred progenitor lines were examined for instances of heterozygosity. In BSSS(R)P, 10 loci were heterozygous. Nine of these were heterozygous in single progenitors. Locus *umc26* was heterozygous in 4 of 16 progenitors. In BSCB1(R)P, 39 loci were heterozygous. Locus *bnl13.05* was heterozygous in 3 of 12 progenitors. Heterozygosity of the other 38 loci was restricted to one or two lines. Line K230 was responsible for most of the heterozygosity, as it was heterozygous at 36 loci.

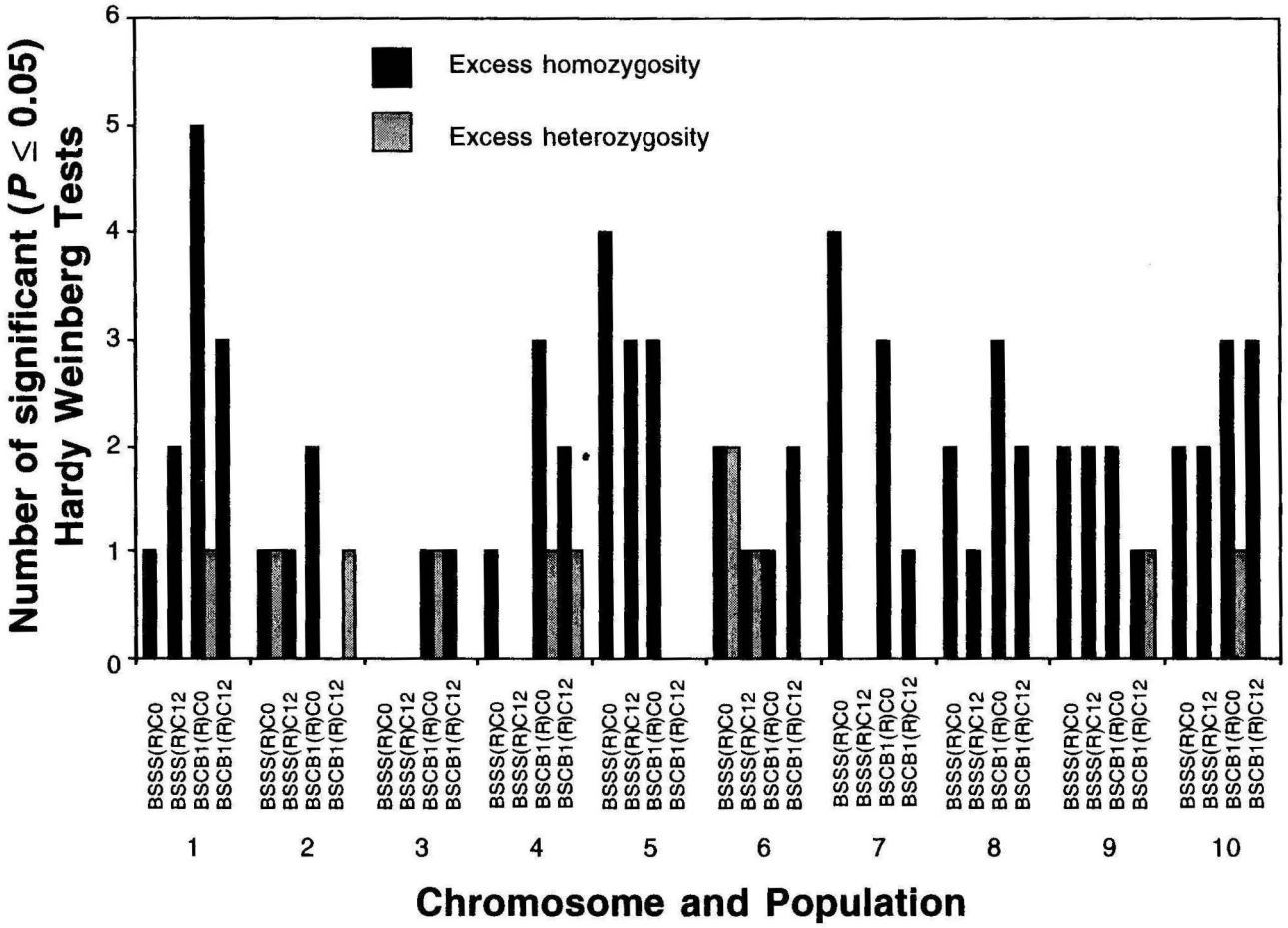


FIGURE 1 - Numbers of significant Hardy-Weinberg tests in two maize populations undergoing reciprocal recurrent selection.

### *Two-locus Equilibrium*

The progenitor populations were essentially in two-locus equilibrium. Less than 0.1%, 1%, or 5% of the tests were statistically significant at the corresponding $P$ levels, which could be explained by type I error (Table 1). BSSS(R)C0 and BSCB1(R)C0 showed increased disequilibrium relative to the progenitors and similar levels of disequilibrium relative to each other. The number of loci involved in disequilibrium in Cycle 0 included every locus at least once at the $0.01 < P < 0.05$ significance level. The matrix in Fig. 3 illustrates the loci arranged spatially based on published maps (see LABATE *et al.*, 1997, for references). If there had been a spatial component to the disequilibrium, i.e. loci on the same chromosome were more likely to be in disequilibrium compared to loci on different chromosomes, more pairs would be found on or near the diagonal of the matrix. This pattern was not strongly evident in any of the six

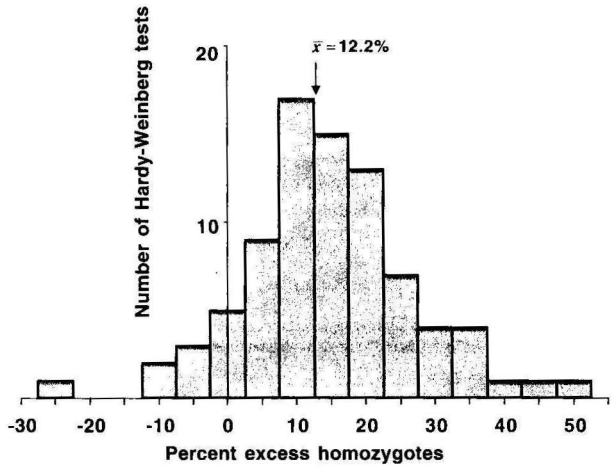


FIGURE 2 - Percent excess homozygotes observed versus number of Hardy-Weinberg tests in BSSS(R)C0, BSSS(R)C12, BSCB1(R)C0 and BSCB1(R)C12 maize populations after random mating for two or more generations.


FIGURE 3A - Pairs of loci in significant linkage disequilibrium with each other for 82 RFLP in BSSS(R) and BSCB1(R) progenitor, Cycle 0 and Cycle 12 maize populations. Loci are sorted by chromosome and map position. No fill, ns; black fill, $P \leq 0.05$.

populations. One visible pattern was that chromosome three showed less disequilibrium than the other chromosomes. This may have been because there were fewer markers on chromosome three.

The two Cycle 12 populations differed dramatically from each other in their levels of disequilibrium. BSCB1(R)C12 showed fewer cases of disequilibrium than BSCB1(R)C0, although the decrease was small when adjusted for the total number of pairs (Table 1). BSSS(R) displayed a vast increase in incidences of two-locus disequilibrium between Cycle 0 and Cycle 12. When the identities of loci in disequilibrium were examined it was noticed that many of the loci near fixation showed a high degree of disequilibrium with each other and with many other loci spread throughout the genome (Fig. 4). The loci in Fig. 4 were sorted based on the frequency of the most common allele. Two or three of the 100 plants sampled were responsible for creating the pattern of disequilibrium between loci near fixation and the remainder of the genome (evident in Fig. 4). In these individuals, a rare allele was found (usually in het-

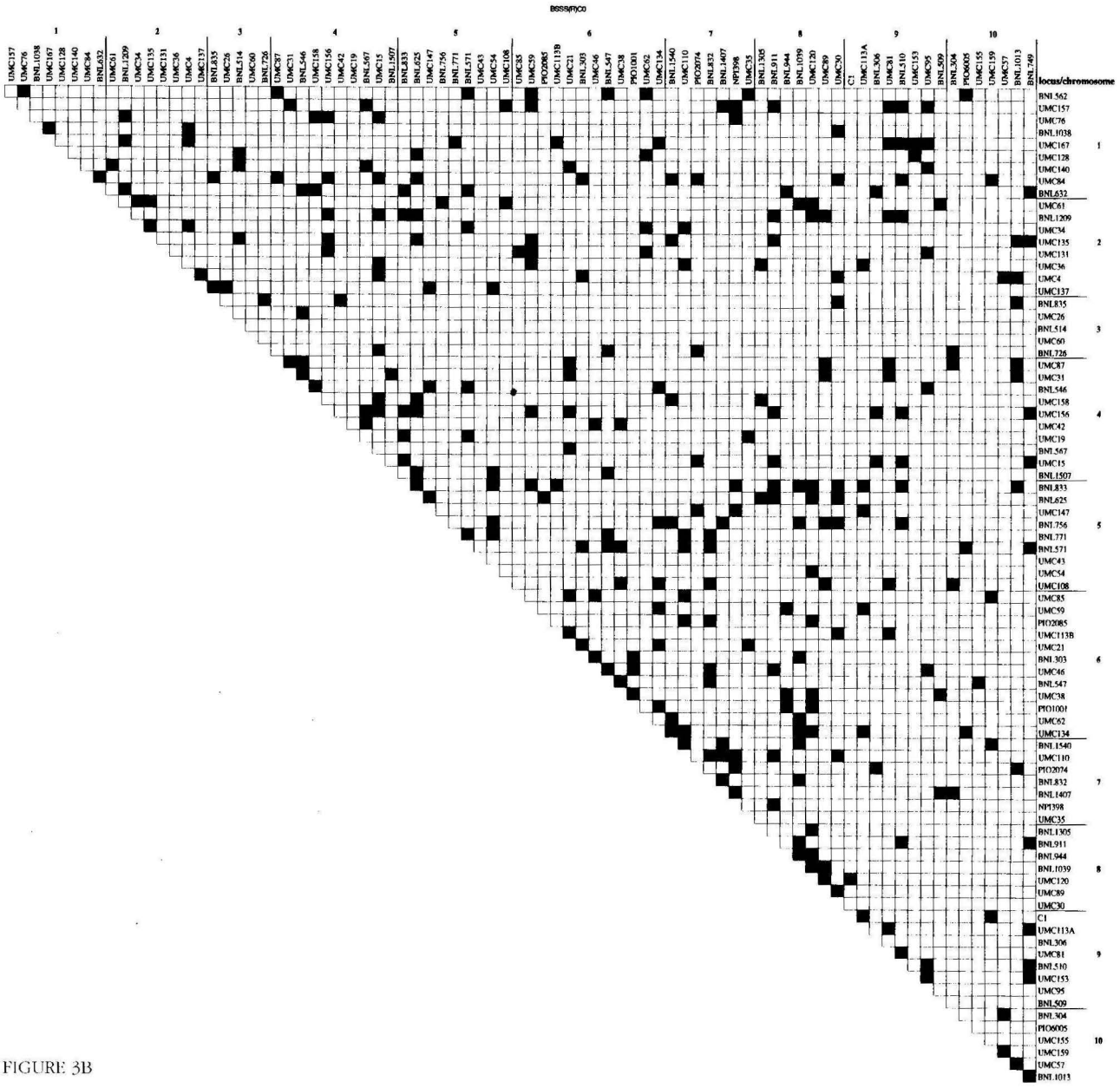


FIGURE 3B

erozygous condition) while the common allele was fixed in every other individual. At loci with the most common allele at relatively low frequency (e.g. *umc84*), a rare allele was present, still restricted to the same two or three individuals.

The PCA further illustrated this phenomenon. With extensive two-locus disequilibrium, the expectation was that two distinct groups of BSSS(R)C12 individuals would be found using PCA. There was no evidence for this except for the three individuals that fell into the Cycle 0 cluster (Fig. 5).

### *PCA*

The results of the PCA (Fig. 5) showed that the two progenitor populations were genetically very similar. BSSS(R)P and BSCB1(R)P did not form two distinct clusters. One BSCB1 progenitor was an extreme outlier, line K230. Previous evidence had suggested that K230 was possibly contaminated before being sampled for our study (LABATE *et al.*, 1997).

In the absence of genetic drift and selection, the Cycle 0 populations should have remained clustered with the progenitors. We have inferred that several

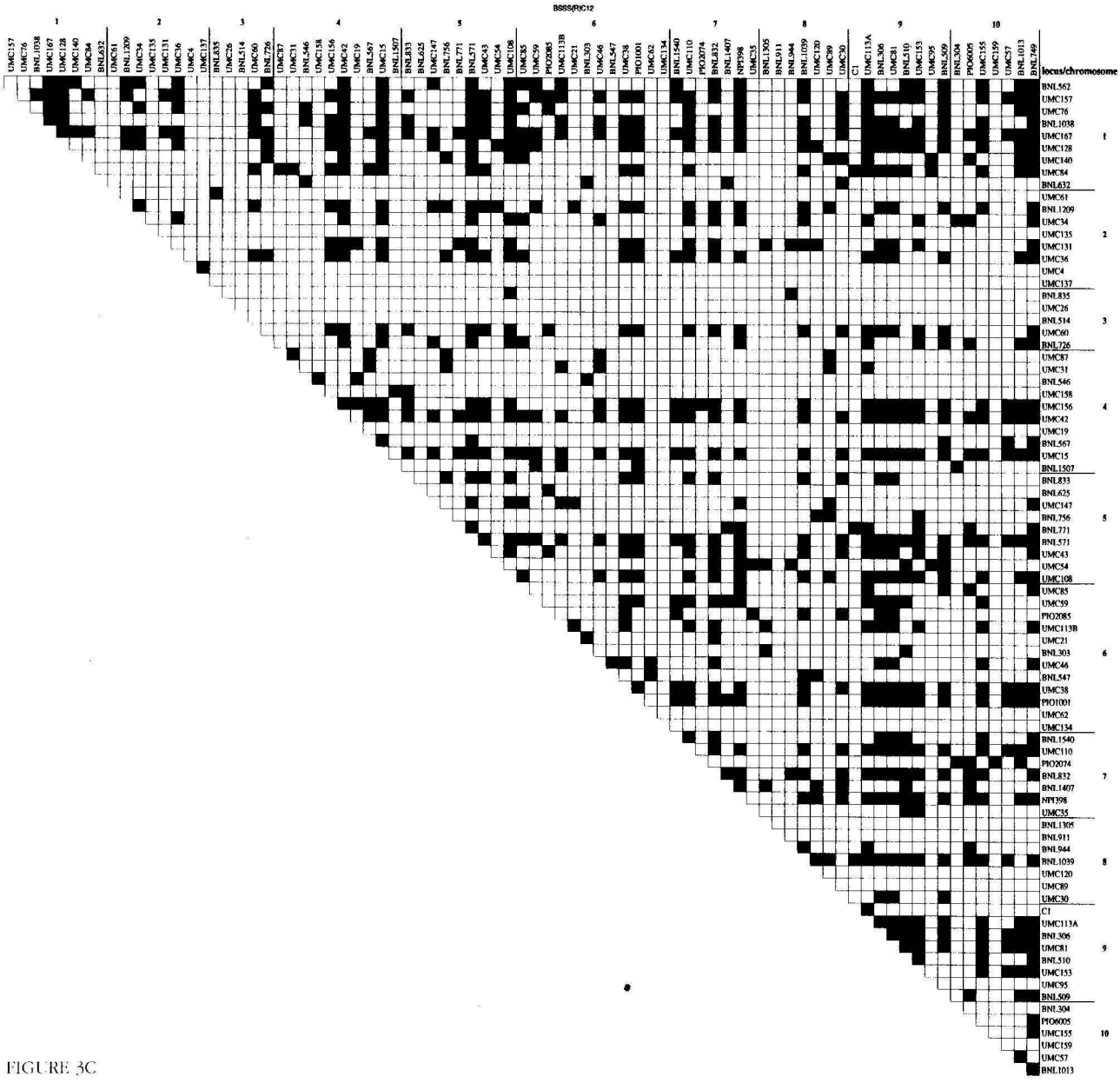


FIGURE 3C

decades of maintenance of BSSS(R)C0 and BSCB1(R)C0 has altered their genetic constitutions. This was especially evident in BSSS(R), where it appeared that many rare alleles present in the progenitors were not sampled in the current representatives of Cycle 0 (LABATE *et al.*, 1997; LABATE *et al.*, 1999). By Cycle 12, BSSS(R) and BSCB1(R) were substantially diverged. A few outliers could be seen for the Cycle 0 and Cycle 12 populations, but in general the clustering was strong. The first three principal components accounted for 78.4% of the total observed variation (61.9%, 11.4%, and 5.1% respectively).

## DISCUSSION

One of the first assumptions made in quantitative genetics is Hardy-Weinberg equilibrium (FALCONER and MACKAY, 1996, p. 5). The assumption of random mating is central to the development and interpretation of quantitative genetic theory. Devia-

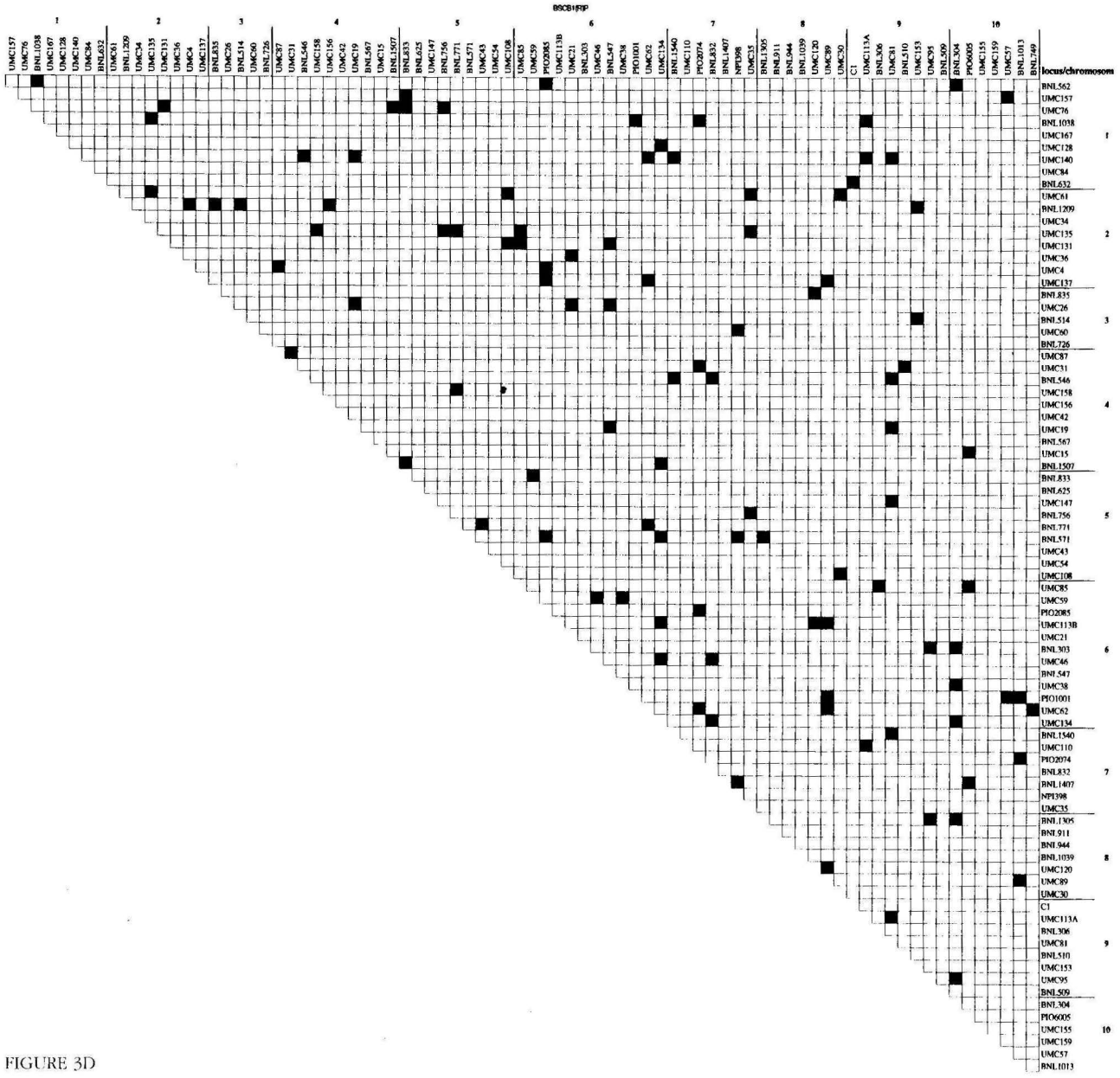


FIGURE 3D

tions from the assumption of random mating can be determined theoretically, however, only a few studies have reported on the empirically estimated equilibrium status of maize populations.

Hardy-Weinberg equilibrium is the maintenance of Hardy-Weinberg proportions (HWP) over successive generations, and implies the absence of disturbing forces such as mutation, migration, and selection as well as the continuation of random mating (WEIR, 1996, p. 95). The Hardy-Weinberg test ascertains whether a population exhibits genotypic frequencies consistent with HWP, but accepting the null hypothesis does not guarantee that the population is free of disturbing forces (HARTL and CLARK, 1989, p. 36-37).

Randomness of mating in maize has been the subject of several empirical studies. For 10 populations, including open-pollinated and synthetic varieties, most showed significant excess homozygosity using RFLP and isozyme markers (DUBREUIL and CHARCOSSET, 1998). This was interpreted as probably resulting from nonrandom mating. In two synthetic

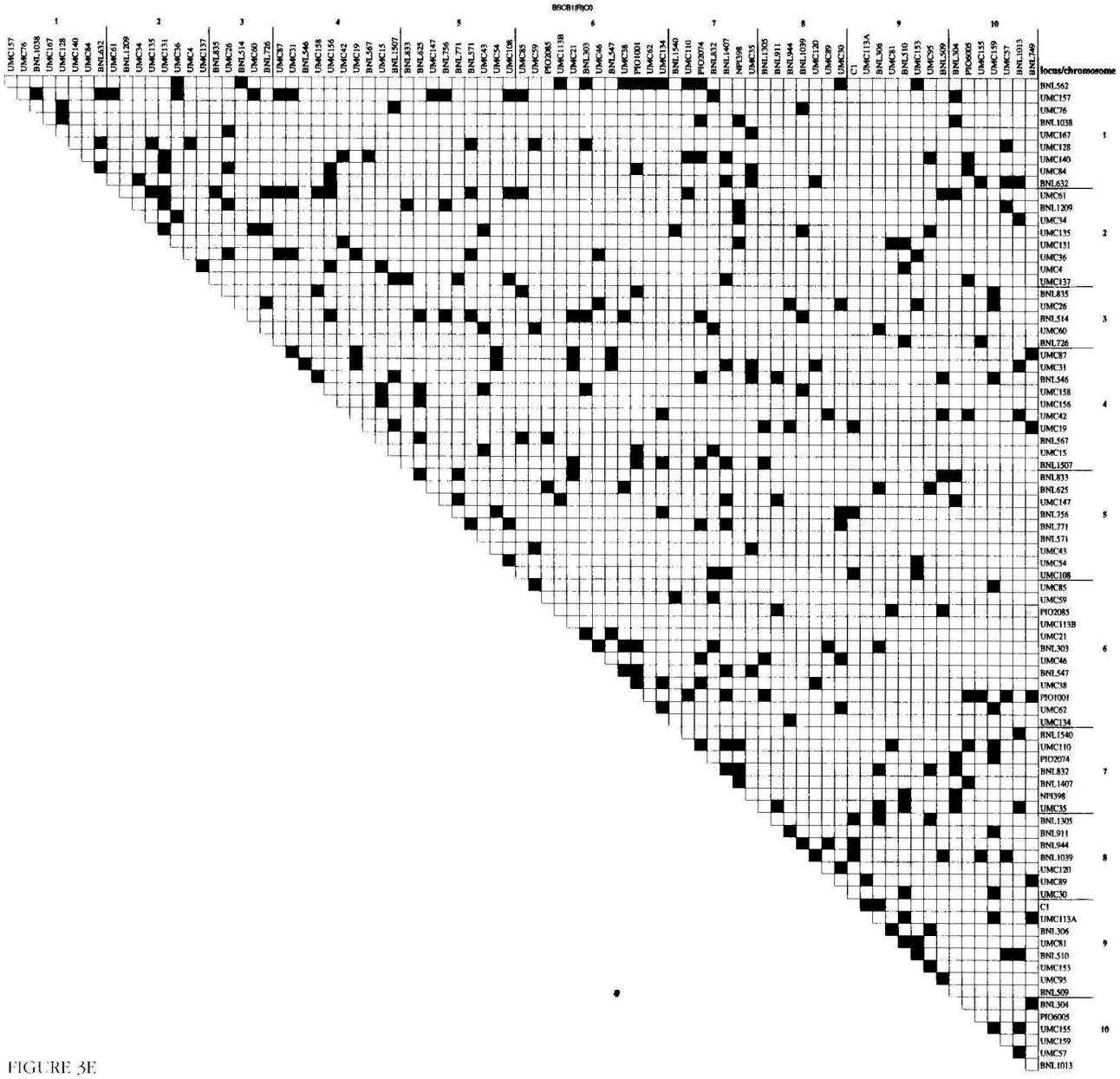


FIGURE 3E

populations undergoing $S_1$ recurrent selection, about 10% of the tests for HWP using allozyne markers were significant, but this was seen as lack of statistical power and results were not reported in detail (REVILLA *et al.*, 1997). In 17 open-pollinated and adapted exotic populations assayed at 13 enzyme marker loci, 27% of Hardy-Weinberg tests were significant, with 94% showing excess homozygosity (KAHLER *et al.*, 1986). Non-random mating and/or natural selection favoring homozygosity were concluded to be common features of these populations. Outcrossing rates in two Hays Golden populations were estimated using isozymes (POLLAK *et al.*, 1984; KAHLER *et al.*, 1984). Outcrossing was the predominant form of mating in the population, the rate of selfing averaged 9 to 10%. This resulted in excess homozygosity immediately after mating, but there was evidence that HWP were restored by the adult stage, possibly due to natural selection favoring heterozygosity. Observed and expected genotype frequencies were reported to be in close agreement for nine enzyme loci assayed in 18 pop-

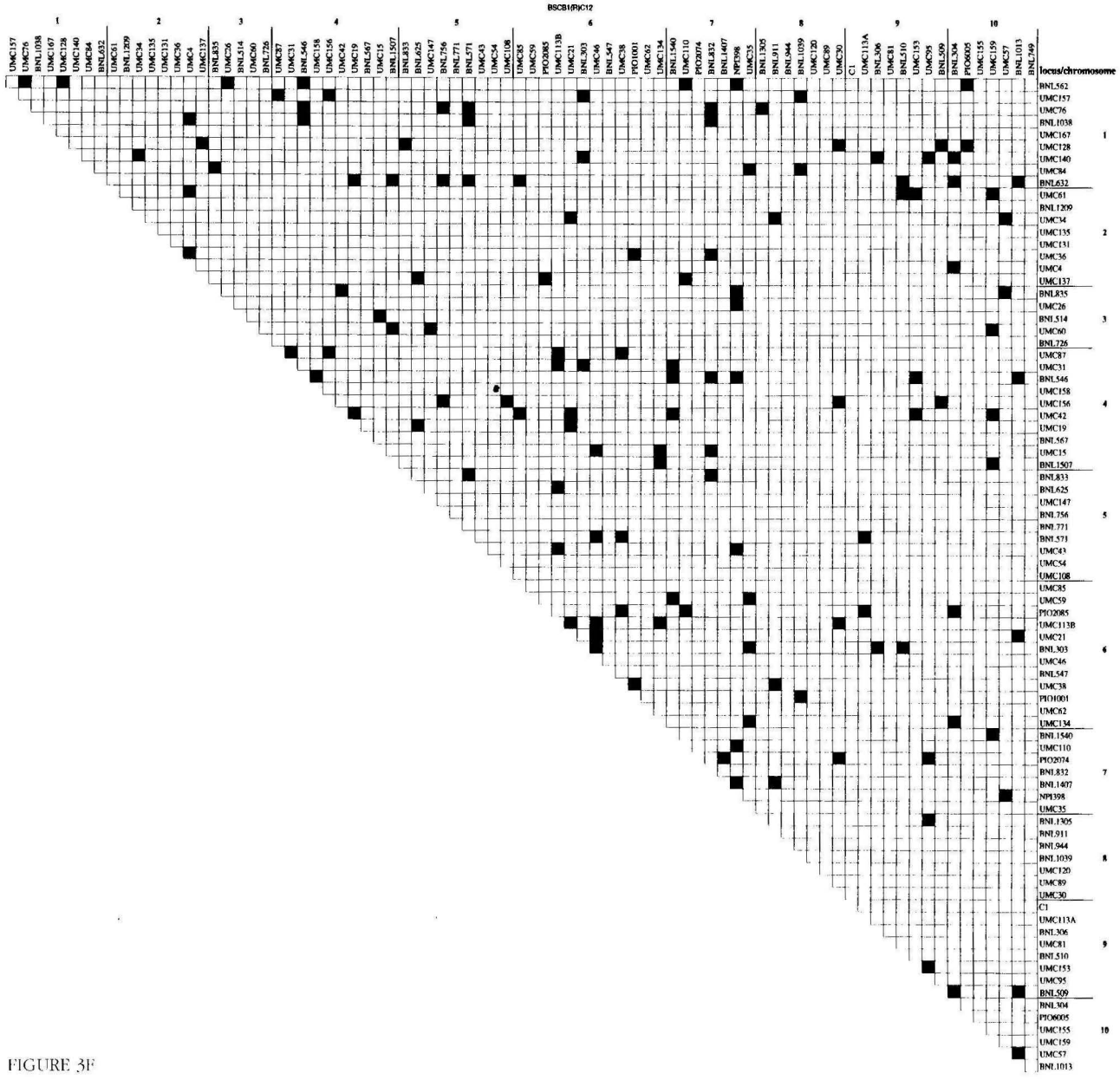


FIGURE 3F

ulations representing eight cycles of Krug BSK half-sib and $S_1$ recurrent selection programs (KAHLER, 1983). In eight Illinois strains under long term directional selection for oil or protein content, six isozyme loci conformed to HWP (BROWN, 1971). Morphological markers were used to examine random mating in isolated polycross blocks (GUTIERREZ and SPRAGUE, 1959). Significant departures from random mating due to an excess of selfing or sibbing were found, possibly related to flowering time.

We detected significant deviations from HWP in 27% of the total of 307 tests performed in BSSS(R)C0, BSSS(R)C12, BSCB1(R)C0, BSCB1(R)C12 populations. A majority of the loci behaved as expected under a model of random mating. The most striking aspect of our findings was the amount of excess homozygosity observed, even at loci that did not depart significantly from HWP (data not shown). Of the 83 instances of significant departures from HWP, 72 or 87% showed excess homozygosity. The mean excess homozygosity for the statistically significant loci over all populations was

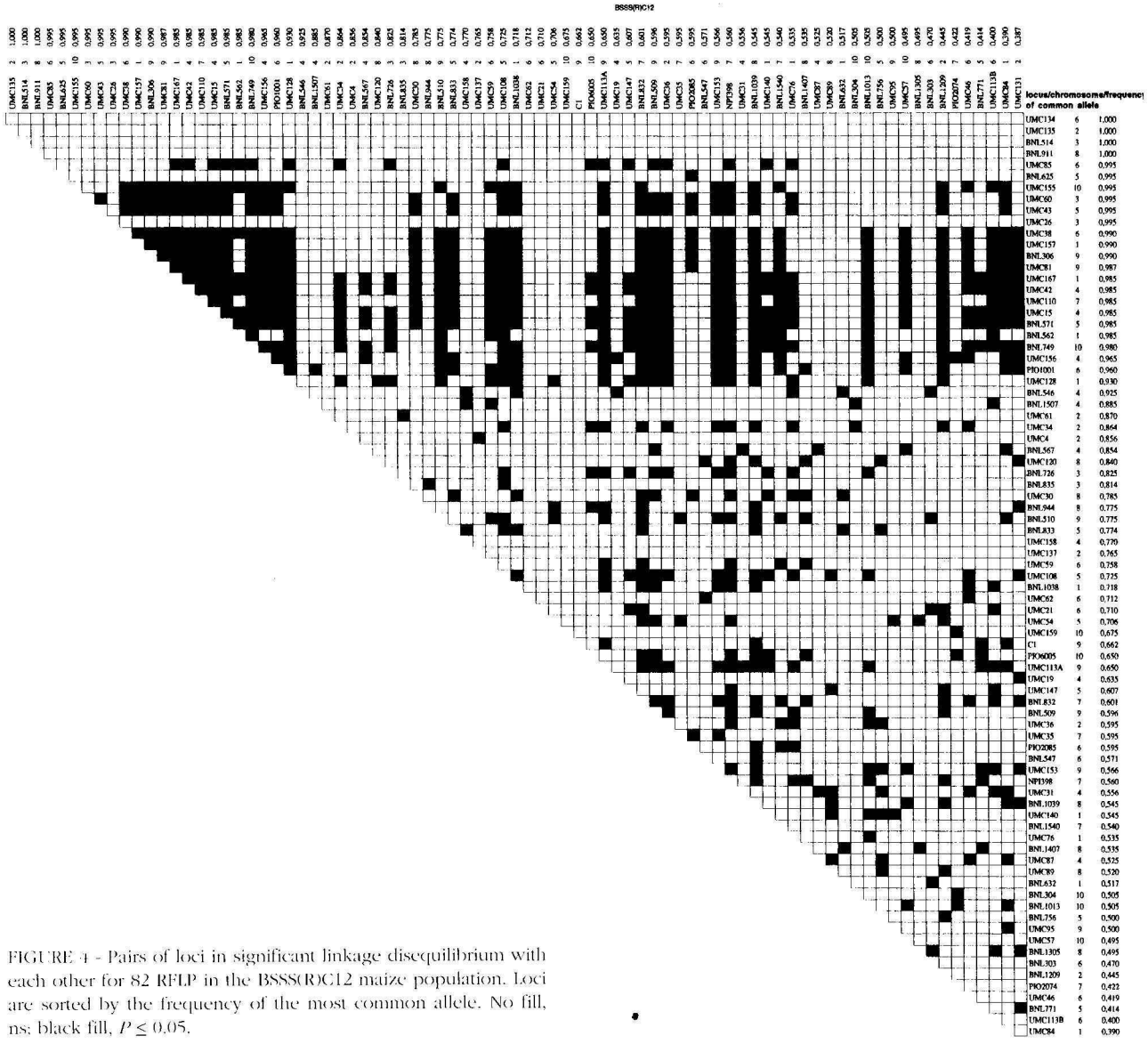


FIGURE 4 - Pairs of loci in significant linkage disequilibrium with each other for 82 RFLP in the BSSS(R)C12 maize population. Loci are sorted by the frequency of the most common allele. No fill, ns; black fill, $P \leq 0.05$.

12.2%, with an expected value of zero if the departures from HWP were due to random causes.

In maize mapping populations excesses of heterozygosity or homozygosity at particular marker loci have been reported (BEAVIS and GRANT, 1991; GARDINER *et al.*, 1993). Locus *umc26* segregated at a ratio of 9:41:6, significantly deviating from the expected 1:2:1 ratio in a Tx303/CO159 immortalized $F_2$ population. This locus was heterozygous in 4 of 16 BSSS(R) inbred progenitor lines. There was only one instance of heterozygosity of *umc26* in the 12 inbred progenitor lines of BSCB1(R), although the same four alleles were present in approximately the same frequencies. Locus *umc26* fit HWP in all four of the Cycle 0 and Cycle 12 populations (data not shown). This could imply natural selection for heterozygosity at this locus in certain genetic backgrounds.

An excess of homozygotes in our BSSS(R) and BSCB1(R) Cycle 0 and Cycle 12 samples can be attributed to one or more of the following: i) the sample sizes used during random mating, ii) positive assortative mating, and iii) the sample sizes used to estimate Hardy-Weinberg equilibrium. Nonrandom components can be introduced into the mating scheme by crossing closely related individuals; this

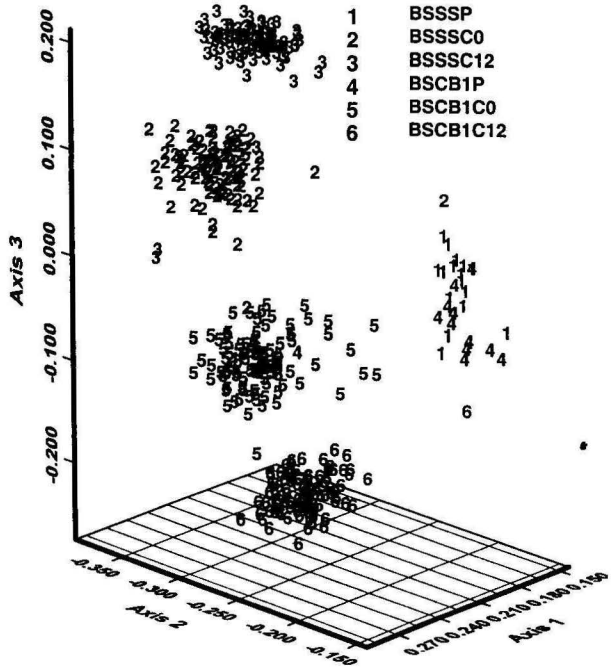


FIGURE 5 - Principal components analysis of BSSS(R) and BSCB1(R) maize populations based on genotypes of sampled individuals at 82 RFLP loci.

can inadvertently occur if excessively small sample sizes are used to advance generations. Sample sizes in this breeding program are approximately 500 representative plants per generation from which 300 to 400 ears are harvested, so this is not likely a contributing factor to excess homozygosity. Positive assortative mating can be caused by genetic variation in flowering time. This is difficult to control, in particular, outliers (early by early and late by late crosses) are suspected as strongly contributing to excess homozygosity in the subsequent generation. This could easily be tested empirically. Sample sizes used to estimate HWP were likely not a factor because sufficient numbers of plants (approximately 100) were sampled from each population. Because the majority of loci in each of the BSSS(R) and BSCB1(R) C0 and C12 were in Hardy-Weinberg equilibrium, the assumption of random mating is generally valid for our breeding program.

Linkage disequilibrium results have been less frequently reported for maize marker studies than Hardy-Weinberg results, although there have been detailed studies in other plants (e.g. BARLEY, see review by HASTINGS, 1990). Linkage disequilibrium may be intra- or intergametic in origin, i.e., correlations exist between alleles at different loci residing on the same gamete or on different gametes. Because we collected diploid rather than haploid data we cannot distinguish between the two. Within a population, the same forces can be responsible for causing deviations from HWP, intragametic, and intergametic (or zygotic) linkage disequilibria (WEIR, 1996, p. 121).

No evidence of an accumulation of linkage disequilibria was observed between eight enzyme loci in four long-term maize selection experiments (STUBER *et al.*, 1980). This included results from nine cycles of RRS in Jarvis and Indian Chief populations. In the Davis RRS program an assay of nine isozyme loci for two cycles revealed an extensive amount of linkage disequilibrium in the progenitor generation, higher than in any of the subsequent generations (BROWN and ALLARD, 1971). The authors concluded that a reduction in population size from combining a small number of parental lines to form the initial populations, and the bottleneck at each cycle that resulted from the selection process, conserved linked blocks of genes and generated chance correlations between unlinked loci. The latter should disappear after one generation of random mating using a large population size.

Because the BSSS(R) and BSCB1(R) progenitor populations were in linkage equilibrium, the disequilibrium that we observed in Cycle 0 and Cycle 12 samples must have been generated over the course of the breeding program. Linkage disequilibrium can be created by i) genetic drift in small populations, ii) hitchhiking of a linked neutral allele with a selected allele, iii) selection for favorable combinations of alleles (epistasis), or iv) migration, or admixture of populations with different gene frequencies. Genetic drift is an obvious candidate for causing the observed linkage disequilibrium in the Cycle 12 populations. The effect of drift on correlations between alleles will increase with decreasing recombination rate and decreasing effective population size ($N_e$) (WOODWARK *et al.*, 1992). Estimates of $N_e$ in BSSS(R)C12 and BSCB1(R)C12 populations are relatively small and they fall within the range of the harmonic mean of the number of selected lines per cycle over 12 Cycles, $N$, to $2N$-1 (12 to 23) (LABATE *et al.*, 1999). However, theoretical studies have shown that drift is unlikely to generate linkage disequilibrium between loosely linked loci (see HASTINGS, 1990). If drift had been the major contributing factor then we would expect to detect linkage disequilibrium between pairs of loci on the same chromosome more often than between pairs of loci on

different chromosomes, but this was not evident (Fig. 3). Furthermore, BSCB1(R)C12 contained approximately equivalent amounts of linkage disequilibrium as BSCB1(R)C0 (Table 1) although the Cycle 12 populations have experienced many more bottlenecks. Linkage disequilibrium increased substantially in BSSS(R)C12 relative to BSSS(R)C0 but this seems likely to have been created through experimental error (see below).

A similar argument can be made for rejecting hitchhiking of neutral alleles with selected alleles as a primary cause of the Cycle 12 disequilibrium. BSSS(R)C12 and BSCB1(R)C12 populations are thought to have experienced extensive hitchhiking effects throughout the genome (LABATE *et al.*, 1999). Hitchhiking creates linkage disequilibrium when a rare neutral allele is tightly linked and correlated with an allele that undergoes a selective sweep (see NEI, 1987, p. 331). Our observations do not fit this scenario because there was little evidence of linkage disequilibrium in the progenitor populations, and unlinked loci account for much of the linkage disequilibrium in the Cycle 12 populations. Selection for epistatic effects must be considered because the Cycle 12 populations are being subjected to intense selection pressure for agronomic performance, and all populations experience natural selection. The amount of linkage disequilibrium did not increase in BSCB1(R)C12 compared to BSCB1(R)C0, therefore, selection for agronomic traits cannot be implicated. The proportion of pairs of loci in linkage disequilibrium increased 2 or 3-fold in the representatives of the Cycle 0 populations relative to the progenitor inbred lines, and this level was approximately the same in the BSCB1(R)C12 population. This could be interpreted as a result of natural selection for epistatic effects between unlinked loci in the maize populations after they were synthesized from the collections of inbred lines.

A critical observation in BSSS(R)C12 is that rare alleles in linkage disequilibrium with each other were involved in the majority of cases of significant linkage disequilibrium, and that these rare alleles were restricted to three individuals that are genetically closely related to BSSS(R)C0 (Fig. 5). This seems an unlikely outcome of artificial or natural selection, because it can simply be interpreted as the absence of genetic divergence of a small fraction of the BSSS(R)C12 population from BSSS(R)C0. Population admixture must therefore be considered, i.e., the possibility that these three samples did not originate from BSSS(R)C12 but were sampled from BSSS(R)C0 and mislabelled. This is the most parsimonious interpretation of the BSSS(R)C12 linkage disequilibrium results and is not impossible. PCA revealed possible misgrouping of individuals from different populations in another maize study (Dubreuil and Charcosset, 1998). Without having carefully examined all of the evidence we may have prematurely concluded that RRS had generated extensive linkage disequilibrium in BSSS(R)C12. We are repeating our analyses using SSR marker loci on newly sampled plants from various cycles of this breeding program. This will allow us to corroborate or reject our present findings. It will be useful to have data on earlier cycles of BSCB1(R) because 18% of the loci were estimated to be at fixation by Cycle 12, making direct comparisons with BSSS(R)C12 difficult.

## REFERENCES


BEAVIS W.D., D. GRANT, 1991 A linkage map based on information from four $F_2$ populations of maize. Theor. Appl. Genet. **82:** 636-644.

BROWN A.H.D., 1971 Isozyme variation under selection in *Zea mays*. Nature **232:** 570.

BROWN A.H.D., R.W. ALLARD, 1971 Effect of reciprocal recurrent selection for yield on isozyme polymorphisms in maize (*Zea mays* L.). Crop Sci. **11:** 888-893.

CROSSA J., 1989 Methodologies for estimating the sample size required for genetic conservation of outbreeding crops. Theor. Appl. Genet. **77:** 153-161.

CROSSA J., C.M. HERNANDEZ, P. BRETTING, S.A. EBERHART, S. TABA, 1993 Statistical genetic considerations for maintaining germplasm collections. Theor. Appl. Genet. **86:** 673-678.

CROSSA J., S. TABA, S.A. EBERHART, P. BRETTING, R. VENCOVSKY, 1994 Practical considerations for maintaining germplasm in maize. Theor. Appl. Genet. **89:** 89-95.

DUBREUIL P., A. CHARCOSSET, 1998 Genetic diversity within and among maize populations: A comparison between isozyme and nuclear RFLP loci. Theor. Appl. Genet. **96:** 577-587.

FALCONER D.S., T.F.C. MACKAY, 1996 Introduction to quantitative genetics. 4th ed. Longman, Essex, England.

GARDINER J.M., E.H. COE, S. MELIA-HANCOCK, D.A. HOISINGTON, S. CHAO, 1993 Development of a core RFLP map in maize using an immortalized $F_2$ population. Genetics **134:** 917-930.

GUO S.W., E.A. THOMPSON, 1992 Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics **48:** 361-372.

GUTIERREZ M.G., G.F. SPRAGUE, 1959 Randomness of mating in isolated polycross plantings of maize. Genetics **44:** 1075-1082.

HALLAUER A.R., 1985 Compendium of recurrent selection methods and their application. Critical Reviews in Plant Sciences **3:** 1-33.

HALLAUER A.R., F.J.B. MIRANDA, 1981 Quantitative genetics in maize breeding. 1st ed. ISU Press, Ames, IA.

HANSON W.D., 1959 The breakup of initial linkage blocks under selected mating systems. Genetics **44:** 857-868.

HARTL D.L., A.G. CLARK, 1989 Principles of population genetics. 2nd ed. Sinauer Associates, Sunderland, MA.

HASTINGS A., 1990 The interaction between selection and linkage in plant populations. pp. 163-180. *In:* A.H.D. Brown et al. (ed.), Plant population genetics, breeding, and genetic resources. Sinauer Associates Inc., Sunderland, MA.

KAHLER A.L., 1983 Effect of half-sib and $S_1$ recurrent selection for increased grain yield on allozyme polymorphisms in maize. Crop Sci. **23:** 572-576.

KAHLER A.L., C.O. GARDNER, R.W. ALLARD, 1984 Nonrandom mating in experimental populations of maize. Crop Sci. **24:** 350-354.

KAHLER A.L., A.R. HALLAUER, C.O. GARDNER, 1986 Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. Theor. Appl. Genet. **72:** 592-601.

LABATE J.A., K.R. LAMKEY, M. LEE, W.L. WOODMAN, 1997 Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. **37:** 416-423.

LABATE J.A., K.R. LAMKEY, M. LEE, W.L. WOODMAN, 1999 Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. **99:** 1166-1178.

LOUIS E.J., E.R. DEMPSTER, 1987 An exact test for Hardy-Weinberg and multiple alleles. Biometrics **43:** 805-811.

Nei M., 1987 Molecular evolutionary genetics. 1st ed. Columbia University Press, NY, NY.

POLLAK L.M., C.O. GARDNER, A.L. KAHLER, M. THOMAS-COMPTON, 1984 Further analysis of the mating system in two mass selected populations of maize. Crop Sci. **24:** 793-796.

RAYMOND M., F. ROUSSET, 1995 GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J. Heredity **86:** 248-249.

REVILLA P., M.I. VALES, R.A. MALVAR, A. ORDÁS, 1997 Allozyme frequencies, heterozygosity and genetic distances following $S_1$ recurrent selection in two synthetic maize populations. Theor. Appl. Genet. **95:** 1057-1061.

ROHLF F.J., 1994 NTSYS-pc: Numerical taxonomy and multivariate analysis system. Version 1.80. Exeter Software., Setauket, NY.

SHAPIRO S.S., M.B. WILK, 1965 An analysis of variance test for normality (complete samples). Biometrika **52:** 591-611.

SNEATH P.H.A., R.R. SOKAL, 1973 Numerical taxonomy. 1st ed. W.H. Freeman, San Francisco, CA.

STUBER C.W., R.H. MOLL, M.M. GOODMAN, H.E. SCHAFFER, B.E. WEIR, 1980 Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genetics **93:** 225-236.

WEIR B.S., 1996 Genetic data analysis II. 2nd ed. Sinauer Associates, Sunderland, MA.

WOODWARK M., D.O.F. SKIBINSKI, R.D. WARD, 1992 A study of interlocus allozyme heterozygosity correlations: Implications for neutral theory. Heredity **69:** 190-198.